Establishing Scientific Facts

Victoria Stodden
Department of Statistics
Columbia University

Setting Time Aright, Copenhagen
September, 2011

The Concept of a Scientific Fact

In *Opus Tertium* (1267) Roger Bacon distinguishes experimental science by:

1. verification of conclusions by direct experiment,
2. discovery of truths unreachable by other approaches,
3. investigation of the secrets of nature, opening us to a **knowledge of past and future**.

- described a repeating cycle of observation, hypothesis, experimentation, and the need for independent verification,
- recorded his experiments (e.g. the nature and cause of the rainbow) in enough detail to permit reproducibility by others.

Inductive Scientific Reasoning

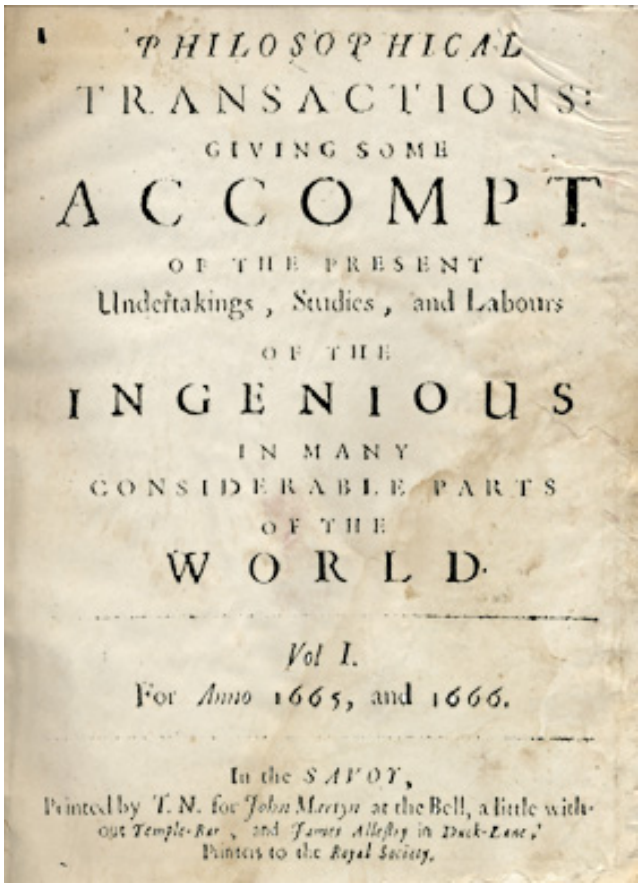
In *Novum Organum* (1620) Francis Bacon proposes:

1. the gathering of facts, by observation or experimentation,
2. verification of general principles.

"There are and can be only two ways of searching into and discovering truth. The one flies from the senses and particulars to the most general axioms, and from these principles, the truth of which it takes for settled and immoveable. ... The other derives axioms from the senses and particulars, rising by a gradual and unbroken ascent, so that it arrives at the most general axioms last of all. This is the true way, but as yet untried."

The Scientific Record

- The Royal Society of London founded 1660 (the "Invisible College"),
- members discussed Francis Bacon's "new science" from 1645,
- Society correspondence reviewed by the first Secretary, Henry Oldenburg,
- Oldenburg became the founder, editor, author, and publisher of *Philosophical Transactions*, launched in 1665.


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.
Vol I.
For Anno 1665, and 1666.
In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.

Scientific Research is Changing

Scientific computation is becoming central to the scientific method:

- Changing how research is conducted in many fields,
- Changing the nature of how we learn about our world.

Conjecture: Today's academic scientist probably has more in common with a large corporation's information technology manager than with a philosophy or English professor at the same university.

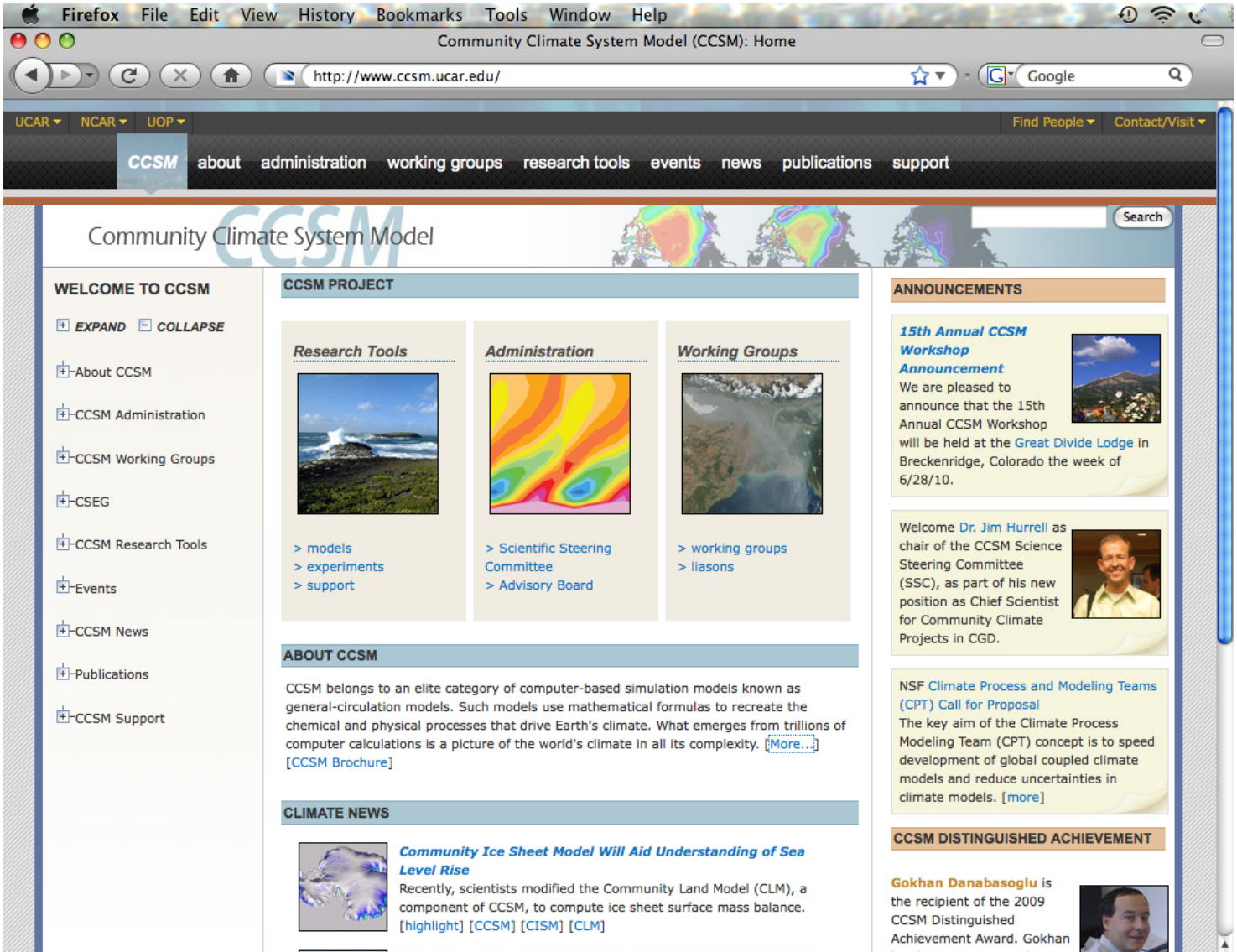
I. Examples of Pervasiveness of Computational Methods

- For example, in statistics:

JASA June	Computational Articles	Code Publicly Available
1996	9 of 20	0%
2006	33 of 35	9%
2009	32 of 32	16%
2011	29 of 29	21%

- Social network data and the quantitative revolution in social science (Lazier et al. 2009);
- Computation reaches into traditionally nonquantitative fields: e.g. Wordhoard project at Northwestern examining word distributions by Shakespearian play.

1. Climate Simulation: Community Climate Models



2. High Energy Physics: Large Hadron Collider

- 4 LHC experiments at CERN: 15 petabytes produced annually
- Data shared through grid to mobilize computing power
- Director-General of CERN (Heuer): "Ten or 20 years ago we might have been able to repeat an experiment. They were simpler, cheaper and on a smaller scale. Today that is not the case. So if we need to re-evaluate the data we collect to test a new theory, or adjust it to a new development, we are going to have to be able reuse it. That means we are going to need to save it as open data." *Computer Weekly*, August 6, 2008

3. Dynamic modeling of macromolecules: SaliLab UCSF

COCEBI-649; NO OF PAGES 12 ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Cell Biology

ELSEVIER

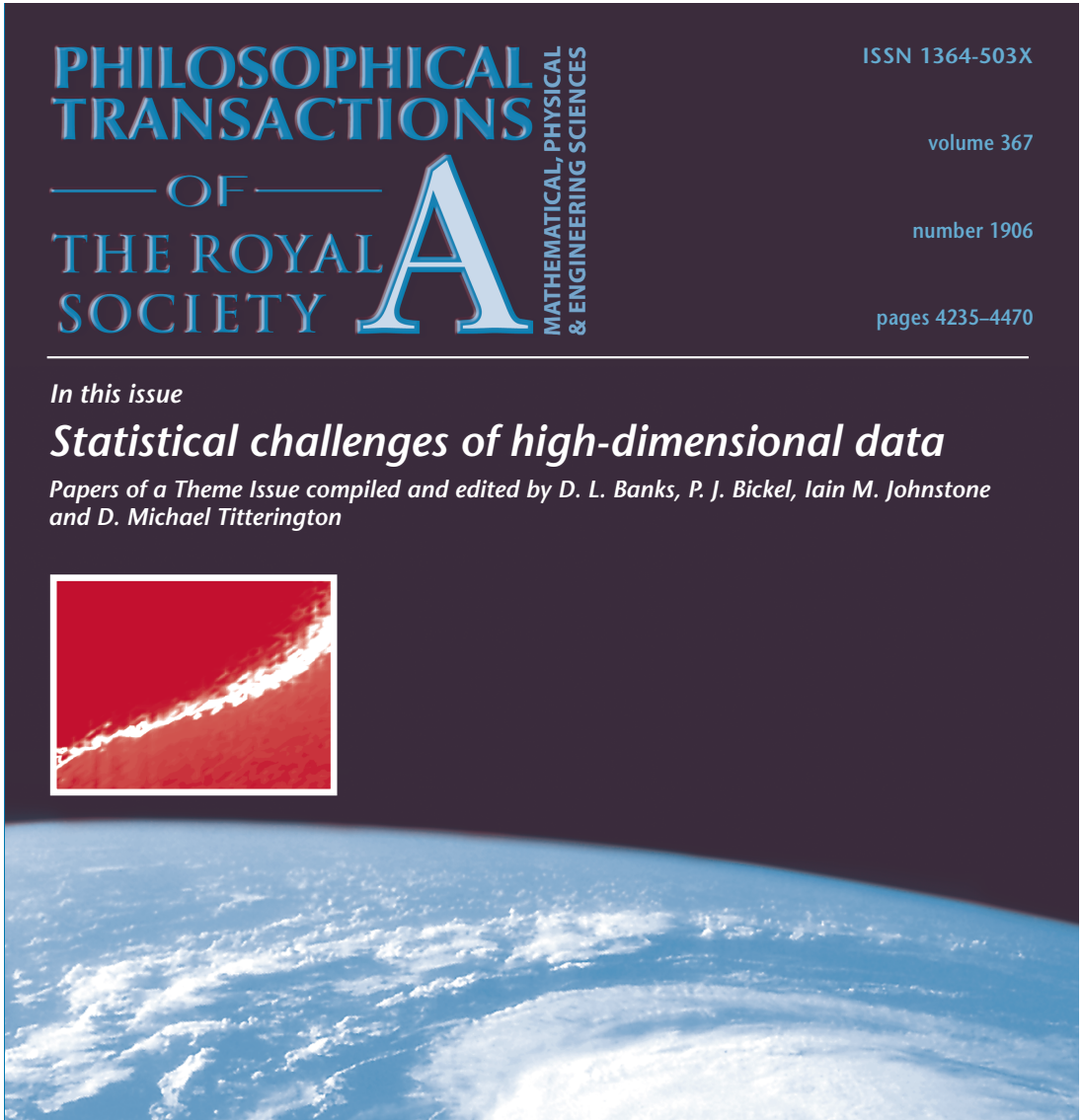
The structural dynamics of macromolecular processes

Daniel Russel[1], Keren Lasker[1,2], Jeremy Phillips[1,3], Dina Schneidman-Duhovny[1], Javier A Velázquez-Muriel[1] and Andrej Sali[1]

Dynamic processes involving macromolecular complexes are essential to cell function. These processes take place over a wide variety of length scales from nanometers to micrometers, and over time scales from nanoseconds to minutes. As a result, information from a variety of different experimental and computational approaches is required. We review the relevant sources of information and introduce a framework for integrating the data to produce representations of dynamic processes.

No single technique, computational or experimental, is able to span all relevant spatial and temporal scales (Figure 3). For static complexes, for example, X-ray crystallography can generate atomic structures of the components, while single particle cryo-electron microscopy (cryo-EM) can provide average mass density maps of the whole assembly at nanometer resolution for the whole assembly. For processes, computer simulations are beginning to reach the microsecond time scale, while

4. Mathematical "proof" by simulation and grid search



Evidence of a problem..

Relaxed practices regarding the communication of computational details is creating a **credibility crisis** *in computational science*, not only among scientists, but as a basis for policy decisions and in the public mind.

Recent prominent examples,

- Climategate 2009,
- Microarray-based clinical trials recently terminated at Duke University.

Clinical trials based on flawed genomic studies

Timeline:

- Potti et al (2006), Nature Medicine; (2006) NEJM; (2007) Lancet Oncology; (2007) Journal of Clinical Oncology: evidence of genomic signatures to guide use of chemotheraputics (*all since retracted*),
- Coombes, Wang, Baggerly at M.D. Anderson Cancer Center cannot replicate, and find simple flaws: genes misaligned by one row, column labels flipped, genes repeated and missing from analysis..
- 2007 correspondence and a supplementary report submitted to the Journal of Clinical Oncology and publication declined; 2008 Nature Medicine declines their correspondence.
- Clinical trials initiated in 2007 (Duke), 2008 (Moffitt).

Clinical trials based on flawed genomic studies

- Duke launches internal investigation Sept 2009; all three trials suspended in Oct 2009,
- Oct 2009: results reported validated, regardless of errors, because data blinded (later found not to be true),
- Jan 2010: Duke clinical trials resume, patients allocated to treatment and control groups. "Neither the review nor the raw data are being made available at this time."
- July 2010: 33 prominent biostatisticians write to Varmus as head of IOM urging suspension of the trials and an examination of standards of review, including reproducibility.
- Sept 2010: IOM committee "Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials" formed,
- Nov 2010: Potti resigns and the clinical trials are terminated.

Controlling Error is Central to Scientific Progress

"The scientific method's central motivation is the ubiquity of error - the awareness that mistakes and self-delusion can creep in absolutely anywhere and that the scientist's effort is primarily expended in recognizing and rooting out error."
David Donoho et al. (2009)

The Third Branch of the Scientific Method

- Branch 1: Deductive/Theory: e.g. mathematics; logic,
- Branch 2: Inductive/Empirical: e.g. the machinery of hypothesis testing; statistical analysis of controlled experiments,

- Branch 3? Large scale extrapolation and prediction, using simulation and other data-intensive methods.

Toward a Resolution of the Credibility Crisis

- Typical scientific communication doesn't include sufficient detail for reproducibility ie. the code and data that generated the findings.
- Most published computational scientific results today are near impossible to replicate.

Thesis: Computational science cannot be elevated to a third branch of the scientific method until it generates *routinely verifiable knowledge*. (Donoho, Stodden, et al. 2009)

Sharing of underlying code and data is a necessary part of this solution, enabling *Reproducible Research*.

Survey of Machine Learning Community (Stodden 2010)

Question: Why isn't reproducibility practiced more widely?
Answer builds on literature of free revealing and open innovation in industry, and the sociology of science.

- Sample: American academics registered at the Machine Learning conference NIPS.
- Respondents: 134 responses from 593 requests ($\sim$23%).

Top Reasons Not to Share

Code		Data
77%	Time to document and clean up	54%
52%	Dealing with questions from users	34%
44%	Not receiving attribution	42%
40%	Possibility of patents	-
34%	Legal barriers (ie. copyright)	41%
-	Time to verify release with admin	38%
30%	Potential loss of future publications	35%
30%	Competitors may get an advantage	33%
20%	Web/Disk space limitations	29%

"Behind one door is tenure - behind the other is flipping burgers at McDonald's."

Top Reasons to Share

Code		Data
91%	Encourage scientific advancement	81%
90%	Encourage sharing in others	79%
86%	Be a good community member	79%
82%	Set a standard for the field	76%
85%	Improve the caliber of research	74%
81%	Get others to work on the problem	79%
85%	Increase in publicity	73%
78%	Opportunity for feedback	71%
71%	Finding collaborators	71%

Grassroots Efforts in Many Fields, Policies

Independent efforts by researchers:

- AMP 2011 "Reproducible Research: Tools and Strategies for Scientific Computing"
- AMP / ICIAM 2011 "Community Forum on Reproducible Research Policies"
- SIAM Geosciences 2011 "Reproducible and Open Source Software in the Geosciences"
- ENAR International Biometric Society 2011: Panel on Reproducible Research
- AAAS 2011: "The Digitization of Science: Reproducibility and Interdisciplinary Knowledge Transfer"
- SIAM CSE 2011: "Verifiable, Reproducible Computational Science"
- Yale 2009: Roundtable on Data and Code Sharing in the Computational Sciences
- ACM SIGMOD conferences
- ...

Policy changes:

- NSF/OCI report on Grand Challenge Communities (Dec 2010)
- NSF report "Changing the Conduct of Science in the Information Age" (Aug 2011)
- IOM "Review of Omics-based Tests for Predicting Patient Outcomes in Clinical Trials"
- NIH, NSF multiple requests for input on data policies
- Journal policy movement toward code and data requirements (ie. *Science* Feb 2011)
- ...

Popular Press

- "The Truth Wears Off," *New Yorker Magazine*, Dec 2010: asserts the 'discovery' of a mysterious effect by which replicated experiments decrease in significance level.
- "it appears that nature often gives us different answers"
- evidence provided in the article:
  - tests on three schizophrenia drugs,
  - Professor Schooler's inability to replicate his own research results,
  - his colleagues' assurances that this happens 'all this time,'
  - ESP experiments from the 1930's,
  - tests for symmetry in sex selection,
  - temporal trends in hundreds of ecology papers.

Question: why bias the publication of results towards ones that agree with previously published results? (Merton's proposed *Universalism* scientific norm)

Popular Press

- "Lies, Damned Lies, and Medical Science," *The Atlantic*, Nov 2010.
- profile of the work of John Ioannidis, Stanford University School of Medicine.
  - exposure of bias and flawed statistical reasoning in medical research,
  - decline effect due to initial 'exaggerations' of the results and researcher error,
  - misinterpretation of p-values, artificial lowering of p-values.

Open Questions Regarding Open Data and Code

- Massive codes or datasets, software support, streaming data,
- Tools for ease of implementation ie. data provenance and workflow, ("progress depends on artificial aids becoming so familiar they are regarded as natural" I.J. Good, 1958),
- Taleb Effect - scientific discoveries as (misused) black boxes,
- nefarious uses? public misinterpretation?
- black boxes and opacity in software (why the traditional methods section is inadequate, massive codebases),
- lock-in: calcification of ideas in software?
- independent replication discouraged?
- policy maker engagement: finding support for our norms.